AMENDMENTS TO THE SPECIFICATION:

Please replace paragraph [0025] bridging pages 7 and 8 with the following paragraph:

- - FIGS. 2A-~~2K~~ 2J depict diagrams of several variant constructions of retroviral vectors which perform certain distinct functions for acquiring different types of information in cells. The critical portion is the area located between the 5' and 3' LTR. These expression cassettes would be moved essentially intact between any of the various viruses and/or plasmids that we have mentioned. FIG. 2A is a vector for exon acquisition. FIG. 2B depicts a vector designed for integration site acquisition. FIG. 2C depicts a vector for incorporation of multiple marker genes. FIG. 2D depicts a transfection cassette. FIG. 2E depicts a vector for replication compliant virus. FIG. 2F depicts a vector for a fusion protein marker for cell pre-separation and FACS analysis. RE (Type IIS restriction enzyme site); LTR, (long terminal repeat); CMV IE, (CMV intermediate early promoter); NeoR; (neomycin resistant gene); pA, (bovine growth hormone poly-A signal); SA, (human γ-globin intron #2 splicing acceptor); pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Gag, pol, env, retroviral helper virus. FIG. ~~2H~~ 2G depicts examples of HIV-1 based vectors that do not have selection markers and have the exon trapping marker hrGFP or AcGFP flanked by either a splice acceptor site only (pHSG) or by both splice acceptor and donor signals (pHSGEX, pHS2GEX3, pHS2AcGEX3 and pHS3GEX4) which are respectively flanked by flexible linker/s of glycine and serines (GC linkers) on the amino terminus only (pHSGEX) or on both amino and carboxyl terminus of the markers hrGFP (pHS2GEX3) or AcGFP (pHS2AcGEX3, pHS3AcGEX4). Vector pHS3AcGEX4 contains an ampicillin resistance gene and a pUC bacterial origin of replication, that allows to perform rescue of the genomic sequence tags flanking the 3'LTR chromosomal insertion site by digestion with RsaI and self-ligation of chromosomal DNA followed by transformation into bacteria and selection for Amp resistance. FIG. ~~2I~~ 2H shows HIV-1 based vectors that in addition to the gene trapping exon markers hrGFP or AcGFP flanked by GC linkers, contains an expression cassette for selection of transduced cells. The expression cassettes consist of either a PGK promoter or an adenovirus Elb promoter driving the expression of either murine α(1,3) galactosyl transferase (αGal) or neomycin phosphotransferase (Neo) genes. Expression of αGal results in the

expression of the α-galactosyl epitopes on glycoproteins and glycolipids present on the external cell surface, which allows to magnetically select for cells showing expression of this gene by using antibodies against α-galactosyl residue, which is normally absent in untransduced human cells. The αGal and Neo genes are cloned in a poly-A trapping configuration to select for those insertion events that occur within transcriptional units. Vectors pHS3AcGEX4PA, pHS3AcGEX4E1bA and pHS3AcGEX4E1bNeo contain an ampicillin resistance gene and a pUC bacterial origin of replication that allows rescuing the genomic insertion sites flanking the 3'SIN LTR as described above. FIG. ~~2J~~ 2I shows four vectors that contain a gene trapping exon encoding three copies of the influenza haemaglutinin HA epitopes. Vectors pHS2HA3Xf0, pHS2HA3Xf1 and pHS2HA3Xf2 encode three copies of the HA epitope in translational frames 0, 1 and 2, respectively. Vector pHS2HA3F encodes three copies of the HA epitope, each one in a different translational frame. This allows for magnetic selection of cells showing gene trapping events that result in membrane proteins displaying the HA epitope in the external surface of the cell membrane. FIG. ~~2K~~ 2J depicts examples of vectors based on MoMLV backbones where the translation frame of the fluorescent marker protein is shifted by zero, one or two nucleotides inserted after the splice acceptor signal. More structural features of these and other vectors are described in Table 1. - -

Please replace paragraph [0046] bridging pages 14 and 15 with the following paragraph:

- - FIG. 20 is a diagram **(SEQ ID NO: 4)** demonstrating the layers of information which may be assayed to identify the real state of cell (furthest outward circle). Those who assay DNA and raw sequence data determine gene function based on sequence similarity, gene structure, and evolutionary relationships. Missing from this data is any mRNA or translational modification data. Those who assay mRNA gain a prediction of a protein profile based on the assumption that protein levels are directly proportional to mRNA. An assumption which is proving to be erroneous. Closest of all these methods to the real cell state is the method of the invention which detects actual cellular protein levels by direct measurement. - -

At page 15, please replace paragraph [0047] with the following paragraph:

- - FIG. 21 is a depiction **(SEQ ID NO: 5)** of a successful gene trapping in pGT5A-transfected PA317 cells. NcoI restriction site located at the 5' end of hrGFP marker gene and an EcoRI at the Oligo-dA primer were used as cloning sites for gene trapped sequence into a sequencing vector which was digested with NcoI and EcoRI. After BLAST searching against mouse EST database in GenBank, the sequence trapped by pGT5A demonstrates 99% homology to a high mobility group protein, HMGI-C, a nuclear phosphoprotein that contains three short DNA-binding domains (AT-hooks) and a highly acidic C-terminus. - -

At page 40, please replace paragraph [0137] with the following paragraph:

- - The preferred embodiment of the invention will use vectors (DNA, RNA, DNA/RNA hybrids etc.) that contain markers which may be sorted to include but not limited to cell surface displayed or cytoplasmic protein; lipid, lipoprotein, glycolipid, and glycoprotein targets that can be tagged with specific fluorescent, chemiluminescent, or bioluminescent compounds using labeled antibodies, direct chemical linkage and/or combination of direct and indirect tagging. These vectors (see FIG. 2A-~~K~~ J, 13A) use either the processes of illegitimate recombination, homologous recombination, and/or viral vectors to integrate said markers into the genomic DNA of target cells (the integrated vector serves as a molecular bar code). Alu sequences are approximately 300 bp in length and are found on average every 3000 bp in the human genome. Alu or other highly repetitive sequences can be used to induce homologous recombination for insertion of the marker gene. The vectors will be delivered to the target cells via standard gene delivery methods to include but not limited to lipid mediated transfection (cationic, anionic, and neutral charged), activated dendrimers (PolyFect™ Reagent, SuperFect™ Reagent {Qiagen}), Phenylethyleneimide (PEI), receptor mediated transfection (fusogenic peptide/protein), calcium phosphate transfection, electroporation, particle bombardment, direct injection of naked-DNA, diethylaminoethyl (DEAE-dextran transfection) etc. Though the preferred embodiment is the use of plasmid based vectors, the use of other high efficiency viral vectors is not precluded. - -

At page 47, please replace paragraph [0160] with the following paragraph:

- - According to the invention serial gene-trapping vectors for the acquisition of the data needed to assign integration sites to specific genes and to mean marker protein expression levels. Examples of such vectors are shown on FIG. 2A-~~2K~~ **2J**. - -

At page 49, please replace paragraph 170 with the following paragraph:

- - In another optional embodiment the expression construct includes a polynucleotide with a negative or positive selection protein for enrichment of the population prior to sorting. Use of the negative or positive selection will remove from the population all cells with no integration of the polynucleotide, for example via antibiotic resistance. This provides for enriched populations of target cells to overcome any relative inefficiency of the gene trapping of genomic control elements. Enrichment of gene trapped cells will include the use of drug selection (e.g., neo^r, $puro^r$, $hygro^r$, zeo^r, HAT^r), affinity separations to include but not limited to Ab/Ag or Ab/hapten, biotin/streptavidin, glutathione S-transferase (GST) fusion proteins, Polyhistamine fusion proteins (Invitrogen), calmodulin-binding peptide tag (Stratagene), HA epitope tag (YPYDVPDYA) **(SEQ ID NO: 1)**, c-myc epitope tag (peptide seq. EQKLISEEDL) **(SEQ ID NO: 2)** (Stratagene), FLAG epitope tag (peptide seq. DYKDDDDK) **(SEQ ID NO: 3)** (Stratagene), V5 epitope (Stratagene), the Linx™. technology {phenyldiboronic acid [PDBA] and salicylhydroxamic acid [SHA] (Invitrogen), adhesion, blocking of adhesion, chemotaxis, block of chemotaxis etc., and/or enrichment by FACS using fluorescent Ab, fluorescent Ag, fluorescent substrates or non-fluorescent substrates that become fluorescent after enzymatic cleavage/activation (A complete listing of common fluorescent probes used for our applications can be found in references: Shapiro, H. M., Practical Flow Cytometry, Third Edition, Wiley-Liss (1994), Robinson, J. P., Handbook of Flow Cytometry Methods, Wiley-Liss (1993); Ormerod, M. G., Flow Cytometry: A Practical Approach, Second Edition, IRL Press (1994); Robinson, J. P., Current Protocols in Cytometry, John Wiley & Sons (2000). - -

At page 86 after the last paragraph of the specification and before the claims, please enter the Sequence Listing enclosed with this preliminary amendment.